Analyzing Significant Locations from Google Timeline Data
Author: Kofi Darfour

Data Sources

The primary data source for this project was the Google Maps Timeline data, extracted directly from the Google Maps mobile application in JSON format. The dataset includes detailed GPS coordinates (latitude and longitude), timestamps marking the start and end of visits, and inferred activities like walking or traveling by vehicle. The data was collected over approximately two and a half months, specifically from January 11th, 2025, to March 29th, 2025. Supplementary information, such as place names and categories, was acquired using the Google Places API. Additionally, route information was fetched using the Google Directions API to visualize precise paths between locations. The full project repository, including the code, visualizations, and data files, is available at: `https://github.com/kodarfour/aobfph_p2`.

Hypotheses

Initial hypotheses for significant locations were based on known frequent visitation points, such as:

- **113-105 Observatory Ave** (Off-Grounds apartment)
- **16581-16473 Hayes Ln** (Family home address)
- **130 Chemistry Dr** (Center of Diversity in Engineering, possibly mislabeled)
- **284-294 McCormick Rd** (Near Clark Hall and Clemons Library)
- **Wilson Hall** (Unexpected frequent visitation, potential GPS clustering issue)
- **Hilton Chicago** (Visited during NSBE National Conference, March 5–9)

Methods

Data Parsing

The JSON file was parsed using Python, extracting location coordinates, start and end times, and duration of visits. Visits shorter than 5 minutes were discarded to eliminate transient GPS readings and short, insignificant stops.

Clustering Algorithm

The DBSCAN clustering algorithm was employed due to its ability to detect clusters of arbitrary shapes and filter out noise effectively. DBSCAN leverages density-based clustering, using parameters `eps` (epsilon) and `min_samples`. The haversine distance metric was selected to account for the curvature of the Earth accurately, essential for geographic data.

The parameters chosen were:

- **eps_meters**: 15 meters — This parameter sets the maximum distance two points can have between them to be considered part of the same cluster. Smaller values produce more precise but potentially fragmented clusters.
- **min_samples**: 2 — Ensures that clusters must have at least two data points, filtering out isolated GPS inaccuracies.

Google Places API

For labeling clusters, the Google Places API was utilized with a radius parameter of 35 meters around the centroid of each cluster to find a matching place name and type. Reducing the radius increases precision but can lead to unlabeled clusters if no location is within the radius.

Route Visualization

The Google Directions API was employed to visualize the routes between clustered locations, providing a more realistic depiction of frequent travel paths compared to direct lines between GPS coordinates.

Results

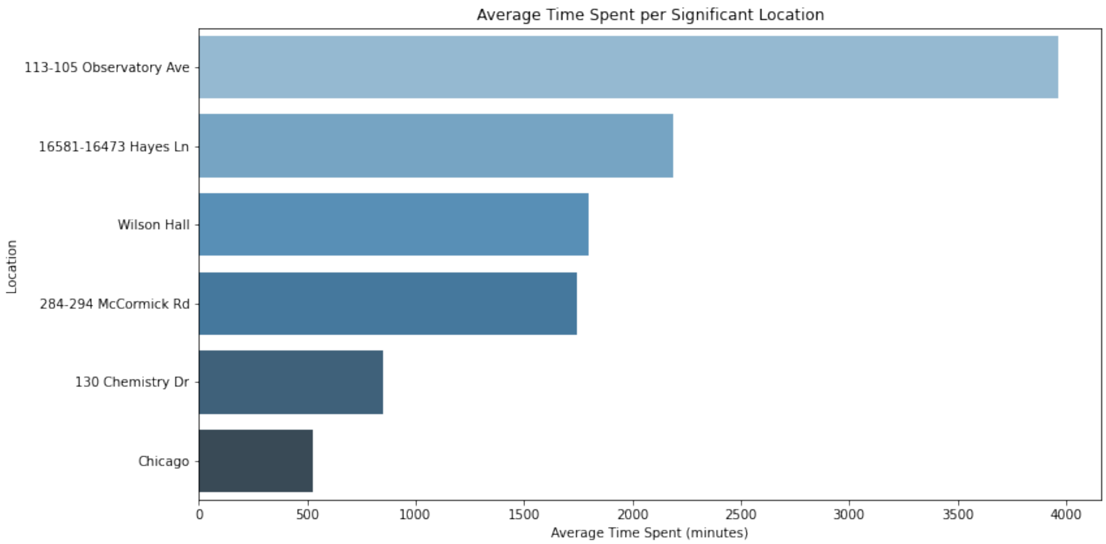

Figure 1: Average Time Spent per Significant Location

The figure above clearly shows the Off-Grounds apartment at **113-105 Observatory Ave** as the most frequented location in terms of average duration. The home address, **16581-16473 Hayes Ln**, and locations around the campus such as **Wilson Hall** and **284-294 McCormick Rd** also stand out.

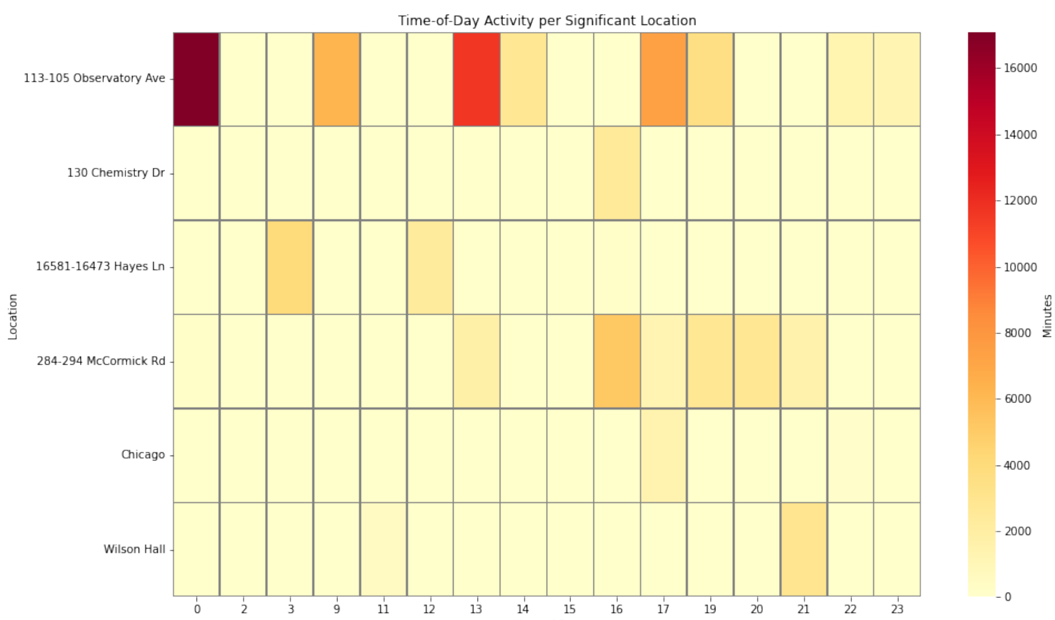

Figure 2: Time-of-Day Activity Heatmap per Significant Location

The heatmap demonstrates clear patterns, especially at the Off-Grounds apartment, with significant activity late at night and in the early morning, reflecting typical residential patterns. Notably, the home address shows less frequent but substantial stays during weekends or breaks.

Interactive Route Visualization

A detailed HTML visualization was generated using Folium, displaying the frequency of visits at various clusters. Each significant location is marked with a circle, whose radius indicates visitation frequency. Routes between locations are visualized, showing the actual paths taken rather than direct straight lines.

For Overleaf embedding, the interactive map is accessible externally through the following link: `https://kodarfour.github.io/aobfph_p2/`

Insights

Several insights emerged from analyzing the data and visualizations:

- The Off-Grounds apartment (**113-105 Observatory Ave**) is clearly the most significant location, reflecting its centrality to daily life.
- The frequent clustering at **284-294 McCormick Rd** likely reflects habitual routes rather than specific destinations, due to proximity to frequently traversed paths near Clark Hall and Clemons Library.
- The identified cluster labeled as **130 Chemistry Dr** is likely a mislabeling or slight GPS error, intended for the Center of Diversity in Engineering, as hypothesized.
- **Wilson Hall** emerged unexpectedly, highlighting the trade-offs of a small radius and precise clustering, likely influenced by passing nearby rather than deliberate visits.
- The location in Chicago (Hilton Chicago) was correctly identified and reflects a unique, temporally isolated event (NSBE Conference).

Trade-offs and Parameter Tuning

The choice of `eps_meters` and search `radius` parameters significantly affects the clustering results:

- A smaller `eps_meters` (15 meters) ensures high precision but may fragment a single significant location into multiple clusters due to minor GPS inaccuracies.
- Increasing the `radius` for Places API queries reduces the chance of unlabeled clusters but increases the risk of incorrect labeling from adjacent locations.

Optimal parameters should be contextually chosen based on the geographical characteristics of analyzed locations (urban vs suburban).

Conclusion

This project successfully leveraged GPS data and clustering techniques to identify and label significant personal locations. The method is broadly applicable for personal analytics, urban planning, and understanding mobility patterns. Further improvements could include adaptive clustering parameters based on urban density and dynamic radius tuning to improve accuracy in labeling clusters.